



<http://digithum.uoc.edu>

La dimensió de les llengües a la Wikipedia i la seua relació amb els elements socials

Borja Pellejero, Natxo Sorolla, Marina Nogué

Xarxa CRUSCAT - IEC

cruscat@demolinguistica.cat

Data de presentació: abril de 2011

Data d'acceptació: abril de 2011

Data de publicació: maig de 2011

Resum

Hi ha un aparent contrasentit en el fet que el català tinga a la Wikipedia un nombre d'articles similar al xinès. Una comunitat que no arriba a deu milions de catalanoparlants, llargament minoritzada al propi territori, pot arribar a tenir una capacitat de producció a internet que en alguns casos és assimilable a la de la Xina, que, amb prop de mil milions de parlants de xinès, és una superpotència econòmica mundial. Els més àvids matisaran que la situació no és la mateixa a la Xina que als territoris de llengua catalana. Aquest text vol fer una primera aproximació a quins són aquests factors socials, educatius, tecnològics, econòmics i demogràfics que estan relacionats amb la posició d'una llengua en el rànquing del nombre d'articles a la Wikipedia. D'aquesta anàlisi naix un concepte clau, el de la *comunitat lingüística digital*, i l'observació que la posició del català en el món d'internet no es deu a un pretès activisme dels seus parlants, sinó més aviat a una posició força semblant a la d'altres comunitats lingüístiques de demografia mitjana.

Paraules clau

Wikipedia, comunitat lingüística digital, català, CRUSCAT, IEC

The number of Wikipedia articles in a language and the links to social factors

Abstract

There would seem to be a contradiction in the fact that Catalan should have a Wikipedia with a similar number of pages to that in Chinese. There are fewer than ten million Catalan speakers, and they were marginalised in their own land for a long time, but they have still been able to produce content on the internet that in some cases matches that of China, a world economic superpower with nearly one billion Chinese speakers. Though it should be noted that the situation is not the same in China as it is in those places where Catalan is spoken. This article offers an initial look at the social, educational, technological, economic and demographic factors linked to a language's position in the ranking of number of Wikipedia articles. This analysis is based on one key concept, that of the digital language community, and the observation that Catalan's position on the internet is not due to the activism of its speakers, but to a position that resembles that of any other medium-sized language community.

Keywords

Wikipedia, digital language community, Catalan, CRUSCAT, IEC


<http://digithum.uoc.edu>

La dimensió de les llengües a la Wikipedia i la seua relació...

Introducció

Les comunitats lingüístiques d'arreu del món no coexisteixen d'una manera aïllada, sinó que estan interconnectades entre elles. Durant els darrers anys han aparegut diversos treballs que han proposat models d'anàlisi que tenen la intenció de reflectir l'ordre lingüístic global (Calvet, 2006; Swann, 2001). Tot i el seu interès, aquests models pràcticament no s'han posat a prova empíricament.

En els darrers anys, aquest sistema lingüístic global ha quedat immers, i possiblement transformat, en el que s'ha conegut com la *societat de la informació*.¹ Internet no és només una nova xarxa de comunicació: també és una xarxa d'interacció, d'informació i d'organització. Internet és la clau d'un nou paradigma de relacions socials a escala internacional. En aquest context, la presència de les llengües a internet té una rellevància cabdal per a l'observació i l'estudi del sistema lingüístic global, i també per a l'anàlisi de la seua evolució en el procés d'expansió internacional.

Els mètodes de mesurament de la posició de les llengües a internet no tenen un consens suficient. Ni tan sols hi ha una eina per a mesurar-ne la difusió. Amb tot, ens podem aproximar a l'objecte d'estudi mesurant la presència d'aquestes llengües en els projectes amb més difusió a internet, que ens poden servir d'indicadors de la presència i extensió de les llengües en aquest àmbit.

Un dels primers punts importants que s'ha de tenir en compte és la diferència entre el consum i la producció de continguts a internet. Segons la posició de l'usuari respecte als continguts, aquest en pot ser consumidor o bé productor.² En l'àmbit de les estadístiques de consum en català, per exemple, ha tingut una certa repercussió la presència del català entre les quaranta llengües amb un volum de consultes més gran al cercador de Google.³

Es tracta d'una dada sobre la presència passiva de la llengua a internet, en un marc en què Google n'acapara les cerques.⁴ Amb tot, fins i tot aquesta gran empresa, amb el volum de dades que maneja, també té alguns problemes per a saber quin és «l'ordre lingüístic mundial» o, si més no, «la llengua» dels seus usuaris. Hi ha dubtes força fonamentats sobre la correspondència entre la llengua preferida per l'usuari i la configuració de la llengua del seu navegador, que en el cas català té conseqüències clares de subrepresentació. A més, les dades de Google no són d'accés lliure ni públic per a fer-ne l'anàlisi.

El camp de la producció de continguts ha adquirit importància, sobretot pel que s'ha conegut com el *web 2.0*.⁵ És un pas basat en la millora de les eines cooperatives en línia, que faciliten la interacció entre els usuaris i la creació de continguts multidireccionals. Un dels màxims exponents d'aquesta evolució d'internet és la Wikipedia,⁶ una enciclopèdia lliure i cooperativa creada pels usuaris.⁷ El projecte ha posat en dificultat el model de negoci de les grans enciclopèdies tradicionals⁸ i se situa en la setena posició en el trànsit general a internet, per darrere de tres cercadors (Google, Yahoo i Windows Live), un espai per a compartir vídeos (YouTube), una plataforma de xarxes socials en línia (Facebook) i un proveïdor de continguts (MSN).⁹

La Wikipedia¹⁰ fou un projecte que començà a funcionar el 15 de gener de 2001. Es caracteritza pel fet de ser una enciclopèdia de llicència lliure,¹¹ en la qual qualsevol persona d'arreu del món que hi estiga interessada en pot editar, copiar i distribuir els articles amb una llicència poc restrictiva. L'obertura del projecte a altres llengües diferents de l'anglès va ser posterior i el català va ser la segona llengua a iniciar-s'hi. Actualment, disposa de més d'11 milions d'articles en 265 idiomes elaborats per més de 150.000 voluntaris.¹² Potser per aquest motiu s'ha convertit en l'enciclopèdia de referència, amb 275 milions de consultes d'articles

1. Castells (2003).

2. Sobre l'ús productiu o de consum a internet, i en especial dels espais col·laboratius, consulteu la bibliografia relacionada amb la teoria del «90-9-1»: el 90% dels usuaris dels webs col·laboratius hi fa consultes, el 9% n'edita els continguts algunes vegades i l'1% hi crea continguts o els edita molt sovint (*Ant's Eye View*, «The 90-9-1 Principle»; *Wikipatterns*, «90-9-1 Theory»).

3. Belzunces (2008), Pozo (2008), *Racó Català* (2008).

4. L'algoritme de Google per a les cerques, innovador en el seu moment, s'ha convertit en omnipresent arreu del món. No per això deixa de tenir competència (*Site analytic*, «Site comparison of google.com, yahoo.com, msn.com»; 1 Cog, «Don't say "search engine" – say "Google"»).

5. *Viquipèdia*, «Web 2.0».

6. Vegeu: <<http://www.wikipedia.org>>.

7. Vegeu: Alexa Internet, «Five-year Traffic Statistics for Wikipedia.org»; Tancer (2007); Woodson (2007); Alexa Internet, «Top 500 Global Sites».

8. Manrique, 2009; Dans, 2009.

9. Alexa Internet, «Top 500 Global Sites».

10. Al llarg del treball fem servir la denominació internacional Wikipedia, que es forma a partir de la paraula *wiki* (sistema web que pot ser editat per diversos usuaris) i el mot *enciclopèdia*. L'origen del nom *wiki* té a veure amb *wiki wiki*, que en hawaïà vol dir «ràpid». En la versió catalana del projecte, s'ha adaptat aquesta denominació a la fonètica del català i s'ha anomenat *Viquipèdia*, tal com també han fet la major part de llengües en les seues respectives edicions (<http://ca.wikipedia.org/wiki/Viquip%C3%A8dia#Noms_alternatius>). Per tal d'evitar la confusió en les adaptacions, respectem la denominació comuna del projecte utilitzant arreu el nom de Wikipedia, i en els casos en què ens referim a una edició en una llengua concreta, també en l'edició en català, acompanyem la denominació amb la llengua d'edició.

11. Vegeu: <http://ca.wikipedia.org/wiki/Viquip%C3%A8dia:Text_de_la_llic%C3%A8ncia_de_Creative_Commons_Reconeixement-Compartir_Igual_3.0_No_adaptada>.

12. Pérez-Lanzac (2009).



<http://digithum.uoc.edu>

La dimensió de les llengües a la Wikipedia i la seua relació...

cada mes, i ha superat les enciclopèdies tradicionals com l'*Encarta* o l'*Enciclopèdia britànica*. A més, aquest projecte enciclopèdic de lliure edició té poca competència en altres projectes de les mateixes característiques. Pareix que només aquest projecte ha obtingut el relleu suficient arreu i és la referència per a totes les llengües.¹³

Així, doncs, a la Wikipedia hi trobem un gran nombre de llengües amb comportaments diferents amb relació a múltiples variables *wikipèdiques*, com el nombre d'articles, d'usuaris i d'edicions, entre moltes altres. Aquesta dimensió de la Wikipedia prové de condicions sociolingüístiques, socials, econòmiques, demogràfiques, educatives o tecnològiques molt diverses. Tenint en compte aquesta composició, creiem que la Wikipedia és un bon camp per a fer una primera exploració de l'ordre lingüístic mundial, almenys el digital.

La recerca sobre els aspectes socials relacionats amb la dimensió de cadascuna de les edicions de la Wikipedia és escadusser. Només una breu repassada dels fons bibliogràfics sobre aquest tema mostra una minsa presència d'aquest tipus de recerca.¹⁴ Els fons actuals reflecteixen la pràctica absència de recerques que s'aproximen a les llengües de les principals edicions de la Wikipedia amb relació a factors socials. Només en algun treball, com el de Van Dijk,¹⁵ s'ha fet una certa recerca quantitativa focalitzada només en les llengües minoritàries. Amb tot, també se centren en factors interns (com el nombre d'usuaris) i en aproximacions de tipus més qualitatiu, i no tant en els factors socials que acompanyen els parlants de cada llengua. Els estudis centrats en els factors interns de la Wikipedia són bastant més nombrosos, però amb un objectiu força diferent al que plantejem en aquest treball.¹⁶

Metodologia

Per a limitar les llengües d'estudi davant la gran quantitat de llengües amb presència a la Wikipedia, el criteri de selecció que s'ha utilitzat és que tinguessen més de 100.000 articles (amb data març de 2009). Hi ha hagut 25 llengües que han complert aquesta condició. Entre aquestes 25, hi trobem llengües amb condicions sociolingüístiques molt diferents: llengües franques o amb un passat colonial important i que es parlen a molts indrets del món, com l'anglès, el francès o el castellà; llengües amb molts parlants i concentrades en un territori molt concret,

com el xinès; llengües mitjanes amb estat propi, com el suec, el finès o el romanès, i sense un estat propi, com el català; i llengües construïdes, en concret, l'esperanto i el volapük. Alhora, aquestes llengües estan inserides en realitats socials diferents. Hi trobem des de llengües que pertanyen a comunitats amb un nivell de desenvolupament tecnològic, educatiu i econòmic elevat, com les llengües nord-europees (finès, suec, danès...), l'anglès o el japonès —que alhora també formen part de contextos culturals diferents—, fins a llengües que pertanyen a comunitats amb un desenvolupament més baix, com el cas de l'indonesi. Entremig, hi ha llengües de països que estan sota la influència soviètica, com el romanès o el rus, o altres com el xinès. Per tant, trobem que és justificat establir com a límit inferior la quantitat de 100.000 articles, que recull una àmplia diversitat lingüística i social.

En aquest treball s'exposa, en primer lloc, la descripció de les variables independents —de tipus social— i de la variable dependent —referent al nombre d'articles de cada llengua a la Wikipedia. Posteriorment es fa una anàlisi de regressió per a trobar relacions entre les variables independents i la variable dependent, amb l'objectiu d'explicar quins són els factors que expliquen millor la dimensió de les llengües a la Wikipedia.

Descripció dels tipus de variables

Entre aquest apartat es presenten les diverses variables socials i sociolingüístiques mitjançant una anàlisi descriptiva univariable. Per aquest motiu, i davant les múltiples variables socials que hi ha, creiem que és convenient separar els diversos factors socials per blocs temàtics, que són els següents: educatiu, tecnològic, econòmic i demogràfic. Primer de tot, explicarem com s'ha gestat el procés de selecció de les variables, les seves fonts i els criteris de ponderació dels valors dels diferents països per a atribuir els usos d'internet a una llengua.

Les variables sociolingüístiques s'han extret principalment de *Linguamón*,¹⁷ i en segon lloc, també d'*Ethnologue*.¹⁸ Les diverses variables socials s'han extret de tres informes de les Nacions Unides: les variables tecnològiques, tret de la inversió en R+D, s'han extret de l'informe *Mesuring the information society* ['Mesurant la societat de la informació'], de l'any 2009, elaborat per la Unió Internacional de Telecomunicacions, que és l'organització més important de les Nacions Unides sobre les tecnologies de la informació i de la comunicació; una part de

13. Fins i tot per a llengües com el castellà, que disposen d'altres projectes similars amb una vitalitat considerable, com és l'*Enciclopedia libre universal en Español* (<<http://enciclopedia.us.es/>>).

14. *Wikimedia*, «Research»; *Wikimedia*, «Wiki Research Bibliography»; *Wikipedia*, «Wikipedia: Academic studies of Wikipedia».

15. Van Dijk (2009).

16. Zlati *et al.* (2006), Ortega Soto (2009).

17. *Linguamón* - Casa de les Llengües. *Les llengües del món*.

18. *Ethnologue*. *Fourteenth Edition*. *Language name index*.



<http://digithum.uoc.edu>

La dimensió de les llengües a la Wikipedia i la seua relació...

les variables demogràfiques s'ha extret de *World Population Prospects*;¹⁹ i la major part de la resta de les variables s'ha extret de *l'Informe sobre el desenvolupament humà 2007-2008*.²⁰

Un cop extretes totes les dades dels diferents països, la qüestió primordial ha estat passar els valors obtinguts, que són accessibles per estats, a cada llengua concreta. A les llengües en què pràcticament la totalitat dels parlants se situen en un estat determinat, se'ls atorga el valor d'aquell estat en concret. Aquests casos són el polonès, el japonès, l'italià, el suec, el noruec, el finès, el català,²¹ l'ucraïnès, el turc, el txec, l'hongarès, l'eslovac i el danès. En canvi, en el cas de les llengües que corresponen a dos o més estats, pel fet que tenen els parlants distribuïts en diverses administracions, s'atribueix a cada estat un valor determinat amb relació a la proporció de la comunitat lingüística concreta respecte dels parlants totals d'aquella llengua. Per tal de poder-ho dur a terme, es van cercar les dades dels parlants de primera llengua en els diferents territoris a partir de les dades de *Linguamón*, en primer lloc, i d'*Ethnologue*, quan la cerca a *Linguamón* era infructuosa. Ara bé, cal comentar dos aspectes rellevants. D'una banda, a les llengües no estatals o que no engloben tot el conjunt de la població d'un estat —com, per exemple, el català a Espanya; l'alemany, l'italià i el francès a Suïssa, i el francès al Canadà—, se'ls atribueix el valor dels diversos indicadors de l'estat. Per exemple, el PIB *per capita* del català és el de l'Estat espanyol, perquè no apareixen les dades regionals, com podrien ser les variables econòmiques de Catalunya, el País Valencià i les Illes Balears, ja que els indicadors comparables no estan disponibles en les estadístiques internacionals. I d'altra banda, s'exclouen de l'anàlisi les dues llengües construïdes, el volapük i l'esperanto, perquè no es poden relacionar amb una comunitat lingüística amb presència territorial i, per tant, no se'n poden obtenir dades.

Les variables sociolingüístiques

En aquest apartat es presenten les diverses variables sociolingüístiques. S'utilitzen tres indicadors que fan referència al nombre de parlants: parlants inicials (L1), parlants de segona llengua (L2) i parlants totals, resultat de la suma d'L1 i L2.²² Cal assenyalar que, en el cas d'algunes llengües, la xifra de parlants no és exacta. Això passa perquè en alguns casos la xifra és única, però en d'altres la dada és un interval de parlants. En aquests casos, s'ha agafat el valor intermedi entre els dos valors de l'interval.

Abans de comentar les tres variables sociolingüístiques seleccionades, cal esmentar que l'ordre d'exposició de les llengües en els gràfics es fa a partir del nombre de parlants com a primera llengua, ja que considerem que ajuda a visualitzar-ho millor. En el gràfic 1, si ens fixem en «Parlants L1», la llengua més parlada amb diferència és la xinesa, amb més de 900 milions de parlants. A molta distància hi trobem el castellà i l'anglès, amb més de 300 milions, seguits del portuguès, el rus i el japonès, amb més de 100. En canvi, llengües com el suec, el noruec, el català, l'eslovè o el danès són de les que tenen menys parlants, amb una xifra que no supera els 10 milions. Finalment, segons les estadístiques, l'esperanto té 1 milió de parlants, mentre que el volapük tan sols té 27 parlants.

Si ens fixem en el nombre de parlants d'L2, veiem que només quatre llengües tenen una xifra significativa en les estadístiques: l'anglès, el francès, el castellà i el rus. Només l'anglès i el francès tenen un nombre de parlants d'L2 superior al d'L1. La llengua que es destaca més és l'anglesa, que arriba quasi a 600 milions de parlants d'L2, mentre que les altres tres se situen prop de 100 milions.

Finalment, si observem el nombre de parlants totals, veiem que l'anglès quasi iguala el xinès amb 900 milions de parlants totals. L'increment de la resta de les llengües respecte a L1 no és tan gran; en la majoria de casos és de prop de 100 milions. D'aquesta manera, el castellà és la tercera llengua més parlada amb quasi 500 milions, seguida del rus, amb 250 (la quarta llengua més parlada) i, finalment, del francès, que no arriba a 200 milions de parlants totals i, per tant, és la sisena llengua més parlada, per sota del portuguès.

Descripció de la Wikipedia

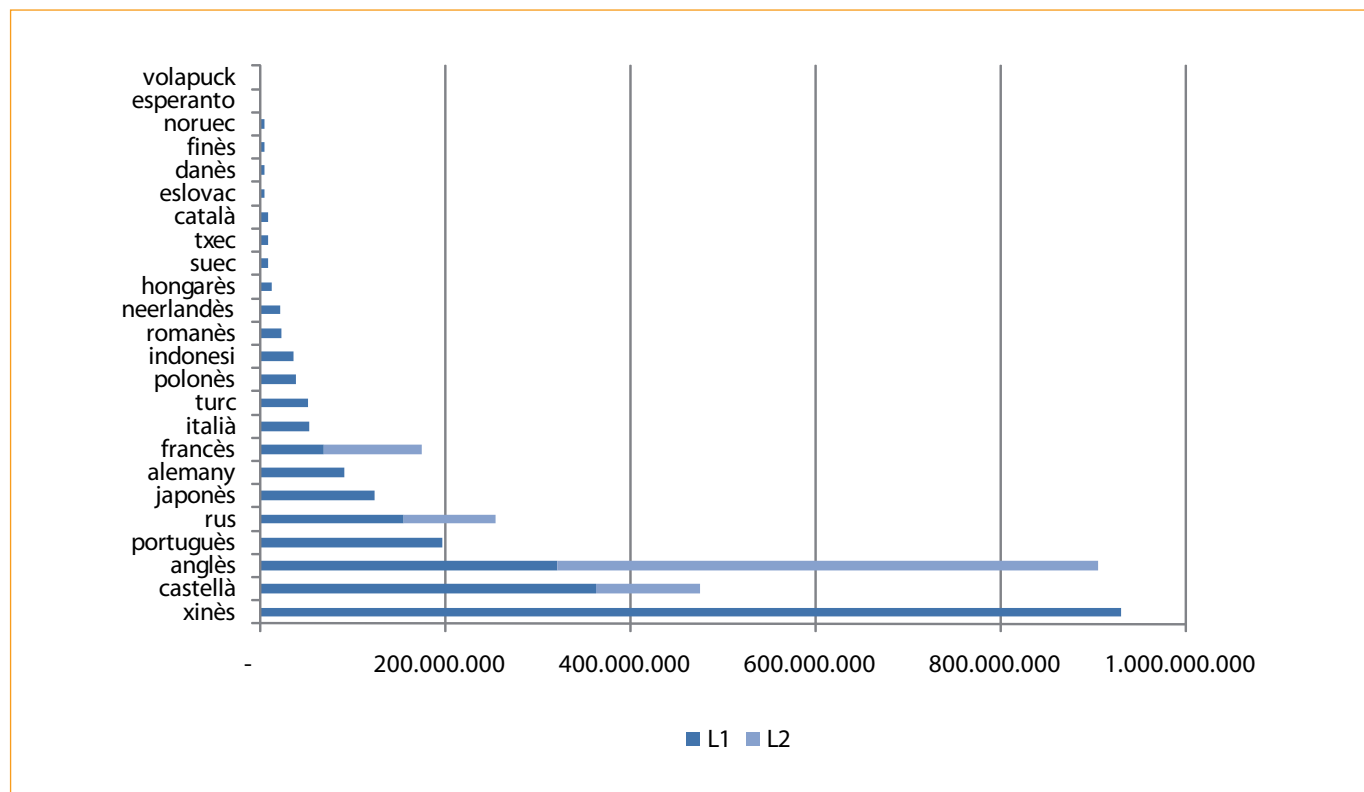
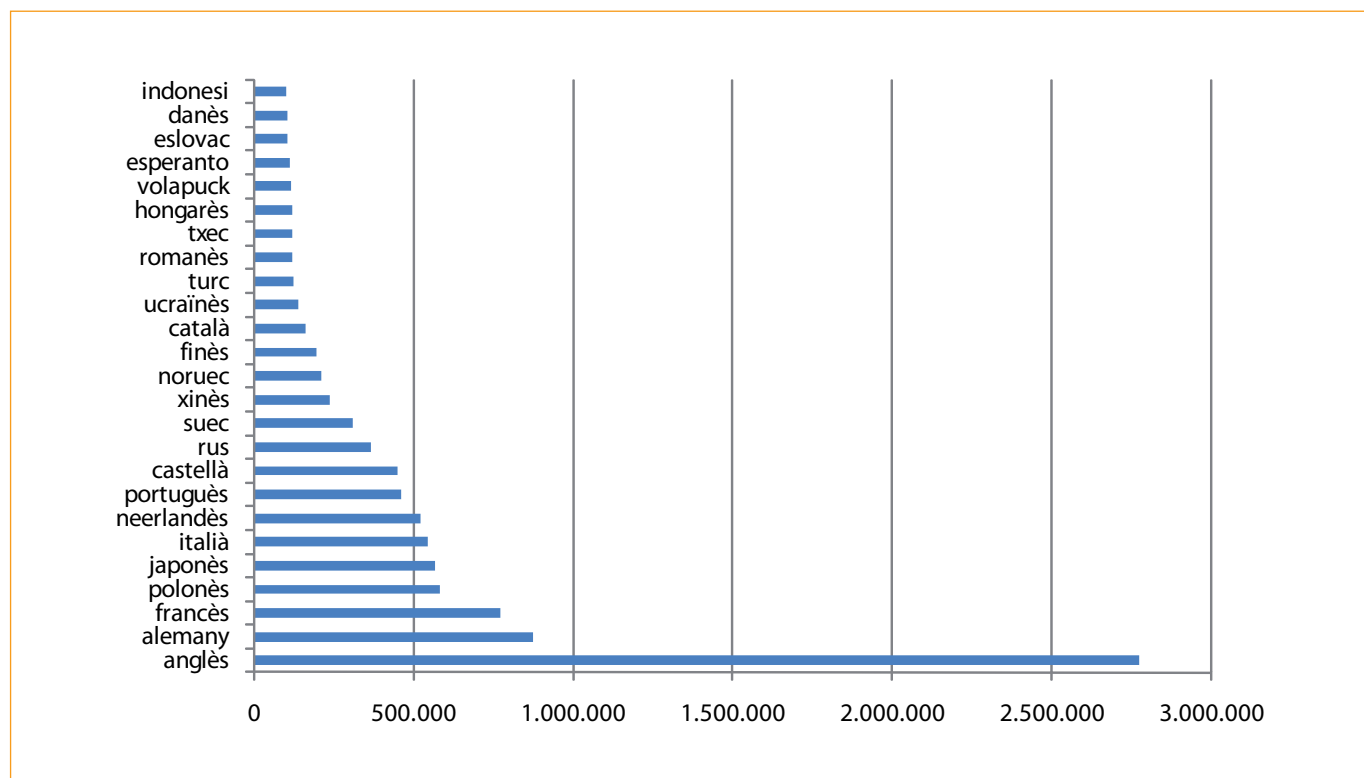
Les dades sobre la Wikipedia són les que s'exposen a la pàgina web *wikimedia.org* i s'han extret a finals de març del 2009. Pel que fa al nombre d'articles a la Wikipedia, l'anglès és la llengua que es destaca més, amb 2.700.000, i obté el triple que la següent, l'alemany, que no arriba a 900.000 articles. Després la segueixen el francès, que no arriba a 800.000; el polonès, el japonès, l'italià i el neerlandès, que en tenen entre 500.000 i 600.000; el portuguès i el castellà, que en tenen entre 400.000 i 500.000; el rus i el suec, entre 300.000 i 400.000; i el xinès, el noruec i el finès, entre 200.000 i 300.000. La resta de llengües se situen entre 100.000 i 200.000.

19. Organització de les Nacions Unides (2008).

20. Programa de les Nacions Unides per al Desenvolupament (PNUD) (2008).

21. En el cas del català, s'han tingut en compte les dades de tot l'Estat espanyol, tant amb dades absolutes com amb dades relatives. No hi ha estadístiques oficials dels territoris de llengua catalana fàcilment comparables a escala internacional, com, per exemple, les de l'ONU.

22. *Linguamón* - Casa de les Llengües. *Les llengües del món*.

**Gràfic 1.** Nombre de parlants d'L1, L2 i parlants totals de les llengües amb més de 100.000 articles a la Wikipedia**Gràfic 2.** Nombre d'articles a la Wikipedia de les llengües amb més de 100.000 articles.



<http://digithum.uoc.edu>

La dimensió de les llengües a la Wikipedia i la seua relació...

Anàlisi causal de les llengües a la Wikipedia

Per tal d'analitzar els principals factors que expliquen la dimensió de cadascuna de les llengües en la Wikipedia, fem una anàlisi de regressió lineal. Aquest mètode estadístic modelitza la relació entre una o diverses variables independents i una de dependent, mesurant-ne la relació causal.

Selecció de variables

El nombre d'articles de cada llengua a la Wikipedia és la variable dependent. La resta de variables introduïdes al model són variables independents, amb les quals intentem explicar el comportament de la variable dependent. Hi ha un gran nombre de variables independents que s'introdueixen al model com a independents i que s'han analitzat en els apartats anteriors: sociolingüístiques, d'educació, tecnològiques, econòmiques i demogràfiques. Mitjançant el mètode de *regressió i introducció* de l'SPSS,²³ se seleccionen les variables que expliquin millor la variació de la variable dependent. Partint d'un model simple, amb la variable independent que s'ajusta millor a la variació de la variable dependent, es construeixen successius que milloren l'anterior mitjançant la introducció d'una nova variable. El procés s'atura en el moment en què la introducció de variables no incorpora la informació suficient i el model perd consistència per un creixement excessiu de la dispersió. Una de les condicions per a introduir una variable és que ha de tenir informació per a totes les llengües, o almenys per a quasi totes. Si una variable no té quantificat un valor per a quasi totes les llengües, en prescindim.

Les variables que finalment s'han introduït són les següents:

- Variable dependent:
 - Nombre d'articles de les 25 llengües principals a la Wikipedia.
- Variables independents:
 - Variables sociolingüístiques: nombre de parlants inicials de la llengua, nombre total de parlants de la llengua (parlants inicials i parlants com a segona llengua).
 - Variables educatives: taxa d'alfabetització de població més gran de 15 anys, taxa d'alfabetització de la població d'entre 15 i 24 anys, despesa pública en ensenyament: proporció del PIB en educació, proporció de la despesa en ensenyament dedicada a primària, proporció de la despesa en ensenyament dedicada a secundària, proporció de la despesa en ensenyament dedicada a l'ensenyament superior, taxa neta de matriculació a primària.

- Variables tecnològiques: índex del desenvolupament de les tecnologies de la informació i del coneixement (IDI), proporció de població usuària d'internet, proporció de llars amb ordinador, proporció de llars amb internet, proporció del PIB dedicat a R+D.
- Variables econòmiques: PIB en dòlars nord-americans, PIB en paritat de poder adquisitiu (PPA), PIB *per capita* en dòlars nord-americans, PIB *per capita* en PPA, taxa de creixement anual de 1990 a 2005, coeficient de Gini: 10% més ric respecte del 10% més pobre, 20% més ric respecte del 20% més pobre.
- Variables demogràfiques: població urbana (%), població jove (%), població adulta (%), població vella (%), taxa anual de creixement demogràfic, taxa bruta de natalitat, taxa bruta de mortalitat, taxa neta de mortalitat, esperança de vida en néixer.
- Índex de desenvolupament humà (IDH).

Un dels primers passos ha estat l'exclusió de l'anàlisi de les llengües construïdes. En aquest cas, parlem de l'esperanto i el volapük. La mateixa naturalesa de les llengües construïdes impedeix que es puguin incloure en aquesta anàlisi. Aquestes llengües es basen en la construcció d'una llengua d'intercomunicació entre parlants inicials d'altres llengües i no tenen una comunitat lingüística consolidada i assentada en un territori determinat. Per tant, no és possible assignar estadístiques de tipus social a les seues reduïdes comunitats lingüístiques i això impossibilita l'anàlisi comparativa amb les altres 23 llengües naturals. A més, el cas del volapük encara és més anòmal, perquè que el gran nombre d'articles existents es deu a la traducció automàtica mitjançant *bots*.

Correlacions entre variables

Un dels principals factors que hem de tenir en compte en les anàlisis posteriors és l'alta correlació entre totes les variables analitzades —sociolingüístiques, educatives, tecnològiques, econòmiques i demogràfiques— i també la variable dependent, que és el nombre d'articles de la llengua a la Wikipedia. La correlació interna per àrees és molt destacada entre les variables pertanyents al camp tecnològic o entre les variables demogràfiques. Si comparem la correlació entre camps, d'una banda, l'alfabetització i la despesa educativa en general van associades amb la difusió tecnològica i amb la renda *per capita*. De l'altra, la pobresa s'associa amb poblacions molt joves i amb un creixement natural elevat. La difusió de la tecnologia s'associa amb la renda *per capita* i, per distorsió dels factors socioeconòmics, amb les poblacions més envellides.

Aquesta correlació entre les variables del model, com veurem més endavant, fa que tot i que es puga jerarquitzar la importància

23. Programari d'anàlisi estadística. Vegeu: <<http://www.spss.com/>>.



<http://digithum.uoc.edu>

La dimensió de les llengües a la Wikipedia i la seua relació...

de les variables a l'hora d'explicar el nombre d'articles a la Wikipedia, s'arriba a un punt on no siga gens senzill aïllar unes variables d'unes altres, i totes en bloc expliquen la variable dependent.

Factors que expliquen el nombre d'articles a la Wikipedia

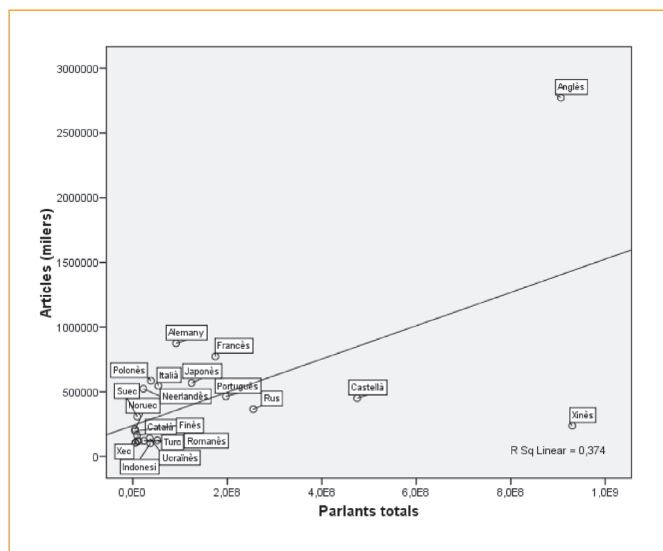
La dimensió de cadascuna de les llengües a la Wikipedia manté una alta correlació amb les variables sociolingüístiques. És a dir, la dimensió de cadascuna de les principals llengües naturals a la Wikipedia està íntimament relacionada amb el nombre de parlants de la llengua. En especial, té una força molt important la suma total de parlants de la llengua, en la qual es compta el nombre de parlants inicials i el de parlants com a segona llengua o llengua estrangera. Aquesta variable, per si sola, s'associa amb la dimensió de la Wikipedia, de tal manera que explica fins al 37,4% de la seua variació.

Per tant, bona part de l'extensió de cadascuna de les llengües a la Wikipedia s'explica per l'extensió de la comunitat lingüística, pel nombre de parlants. Si ens fixem en el gràfic 3, hi ha representades les Wikipedies per uns punts que se situen entre dos eixos, en els quals hi ha el nombre de parlants de la llengua i nombre d'articles a la Wikipedia.

Hi podem observar que els punts es disposen en tres àrees diferents. En primer lloc, a la part dreta superior, hi ha l'anglès, que és la llengua amb un nombre d'articles més elevat (2.773.450) i que se situa en un total de 906 milions de parlants. En segon lloc, hi ha el xinès, que ocupa la part dreta inferior. De la mateixa manera que l'anglès, és una de les llengües amb més pes demogràfic, amb un total de 930 milions de parlants, però el nombre d'articles a la Wikipedia la diferencia de l'anglès: el xinès en té 237.504. Amb un nombre similar de parlants que l'anglès, el xinès només té a la Wikipedia poc més d'una dotzena part dels articles que hi té l'anglès. Aquest fet fa que en la recta de regressió²⁴ dibuixada en el gràfic 3, l'anglès aparega a la part superior i indiqui un nombre d'articles superior al predit pel nombre de parlants, i el xinès aparega per davall, indicant un nombre inferior d'articles.

Finalment, el darrer grup és el de la resta de llengües, que es concentren a la part inferior esquerra del gràfic, en un espai força concentrat. Amb relació a les dues grans llengües, el xinès i l'anglès, són llengües amb menys pes demogràfic. Algunes, com el castellà, tenen un vessant similar al del xinès, amb un nombre d'articles menor del que es predià en la recta. Altres, com l'alemany o el francès, tendeixen a situar-se en la direcció de l'anglès: més articles a la Wikipedia dels que s'han predit pel nombre total de parlants.

Gràfic 3. Model 1. Representació mitjançant un núvol de punts de la dimensió de la Wikipedia segons el total de parlants de la llengua. 23 principals llengües naturals a la Wikipedia. $R^2 = 0,374$.



De la resta, hi ha algunes llengües, com l'italià, el polonès, el neerlandès o el japonès, amb una dimensió més gran que la que s'ha predit, i d'altres, com el rus, el romanès, el turc, l'ucraïnès, el finès, l'indonesi, el txec i, fins i tot, el català, amb una dimensió de Wikipedia més petita que la que s'ha predit pel total de parlants.

En definitiva, veiem que el fet que condiciona el model és una agrupació de llengües (21) amb una dimensió demolingüística diferent i amb un nombre d'articles a la Wikipedia diferent, però que unitàriament contrasta amb el gran nombre de parlants i d'articles de l'anglès. El xinès es configura com un cas atípic, que tot i tenir un nombre de parlants similar al de l'anglès, té un nombre semblant d'articles al de les llengües de potencial demogràfic més petit. En el punt següent veurem com el model intenta resoldre aquesta asintonia amb una nova variable.

En el segon pas per a formar el model, hem cercat la variable que optimitza el model; en aquest cas, el nombre de parlants inicials de la llengua. Novament es tracta d'una variable sociolingüística, però la incorporació d'aquesta variable permet explicar la situació tan distant que hi ha entre el xinès i l'anglès. L'anglès té 906 milions parlants en total, però només una tercera part (321,5 milions) són parlants inicials. La resta de la comunitat lingüística la forma un gruix importantíssim que no la té com a primera llengua. Per contra, pràcticament la totalitat dels parlants de xinès són parlants inicials. La gent que l'ha après com a segona llengua

24. La recta de regressió és aquella que redueix al màxim la distància entre els punts i la línia en un núvol de punts. Això significa que la recta de regressió optimitza millor el tipus de relació entre les variables estudiades. La dispersió entre els punts i la línia és allò que no pot explicar el model.



<http://digithum.uoc.edu>

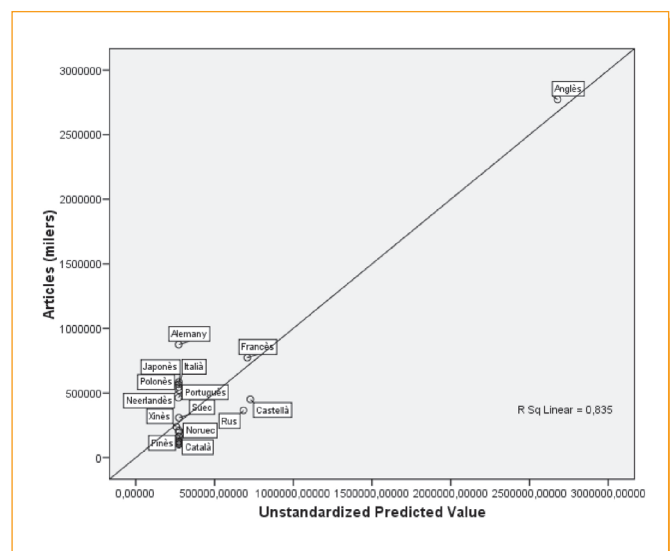
La dimensió de les llengües a la Wikipedia i la seua relació...

o llengua estrangera no és significativa numèricament per a la comunitat lingüística.

La incorporació d'aquesta segona variable es fa en sentit oposat al nombre total de parlants. Així, les llengües que tenen pocs parlants inicials s'impulsen cap a la banda esquerra i les llengües que tenen molts parlants inicials s'impulsen cap a la banda dreta. D'aquesta manera s'aconsegueix que els nombrosos parlants inicials de xinès tendisquen a posar-se a la banda esquerra del gràfic, juntament amb la resta de llengües, mentre que els «pocs» parlants inicials d'anglès queden aïllats a la banda dreta superior del gràfic. Podríem resumir l'aportació del segon model dient que el nombre d'articles a la Wikipedia depèn del total de parlants de cadascuna de les llengües, però aquesta aportació és qualitativament més important en el cas dels parlants no inicials de la llengua, que s'associen amb la capacitat d'atracció de la llengua, qualitativament superior.

Com es veu en el gràfic 4, l'ajustament del model que explica la dimensió de la Wikipedia segons el conjunt de parlants i els parlants inicials és molt superior a l'anterior. Ara s'arriba a explicar el 83,5% de la variació, amb la qual cosa totes les llengües queden ben properes a la recta de regressió. Només l'alemany pareix que queda destacadament més allunyat del valor pronosticat.

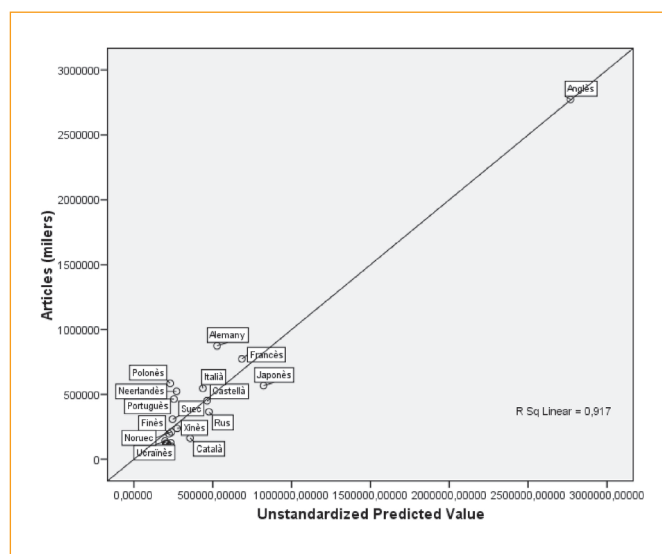
Gràfic 4. Model 2. Representació mitjançant un núvol de punts de la dimensió de la Wikipedia segons la puntuació en la regressió (total de parlants i parlants inicials de la llengua). 23 principals llengües naturals a la Wikipedia. $R^2 = 0,835$.



Per a millorar l'anàlisi i formar el tercer model, s'ha treballat amb el PIB en dòlars nord-americans, que és la variable que aporta la millora més substancial, tot i que la xifra no varia gaire i passa a ajustar-se fins al 91,7% de la variació. Observant el resultat en el núvol de punts del gràfic 5, es percep que la millora es deu,

sobretot, als ajustaments que es produeixen en el nucli de llengües diferents de l'anglès. Si atenem, per exemple, l'alemany, veiem que ha passat a ocupar un espai molt més proper a la recta de regressió, com també ho han fet altres llengües que eren més llunyanes en el model 2, com el castellà, el rus, el japonès o l'italià.

Gràfic 5. Model 3. Representació mitjançant un núvol de punts de la dimensió de la Wikipedia segons la puntuació en la regressió del total de parlants, parlants inicials de la llengua i el PIB. 23 principals llengües naturals a la Wikipedia. $R^2 = 0,917$.



En arribar a aquest tercer model, ja no es pot afegir gaire més informació per cada nova variable que s'hi incorpora. El PIB és una variable força lligada amb el pes demogràfic de la llengua i, per tant, aporta de nou una informació força relacionada, indirectament, amb la dimensió de la comunitat lingüística. A més, són diverses les variables que aporten una quantitat pareguda d'informació a la del PIB i que podrien substituir-la. Aquestes altres variables fan referència a perspectives força diversificades, que no estan tan lligades amb la dimensió de la comunitat lingüística i que aborden camps molt més netament econòmics, tecnològics o d'estructura poblacional: inversió pública en educació (% PIB), desenvolupament tecnològic (IDI), llars amb ordinadors o amb internet (%), proporció de població vella, esperança de vida en naixement o l'índex de desenvolupament humà (IDH).

La conclusió general d'aquestes anàlisis és clara: per a predir l'extensió de les 23 llengües naturals amb més dimensió a la Wikipedia n'hi ha prou de considerar les dues úniques variables de tipus sociolingüístic: el total de parlants de cada llengua i els parlants inicials. La resta de variables que es poden sumar són força més diverses, entre les quals sobreix el producte interior brut. Si es tenen en compte, s'aconsegueix un ajustament de fins al 91,7% del model.



<http://digithum.uoc.edu>

La dimensió de les llengües a la Wikipedia i la seua relació...

També hem observat que la interacció entre la dimensió de cadascuna de les Wikipedies i la resta de variables estudiades la marquen tres directrius. L'anglès, llengua hipercentral en el sistema global i que té una extensió ben patent com a segona llengua, portaria a terme un paper hipercentral també a la Wikipedia i marcaria des de bon començament els requisits bàsics per a ocupar una posició preponderant i, per tant, també els requisits dels models que l'estudien. El xinès, que té una extensa comunitat lingüística, però formada al complet per parlants inicials i, a més a més, amb altres condicionants, tindria una situació important a la Wikipedia, però ni de bon tros un paper hipercentral. La resta de les 21 llengües naturals formen comunitats lingüístiques de mesures inferiors en comparació d'aquests dos casos atípics, i amb això hi aniria associat el nombre d'articles molt inferior als de l'anglès.

Models sense l'anglès i el xinès

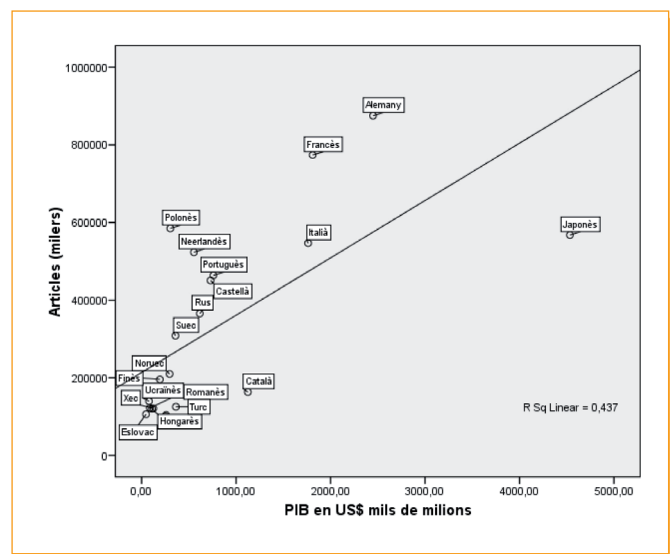
Des d'un bon començament hem exclòs de l'anàlisi les dues llengües construïdes, per la impossibilitat d'analitzar-ne els paràmetres socials des dels qual havíem partit. A més, en el model de l'apartat anterior hem vist el paper atípic que tenen l'anglès i el xinès. Encara que totes dues llengües tenen una comunitat lingüística molt extensa, l'una desenvolupa una dimensió colossal a la Wikipedia i l'altra s'emmotlla en una dimensió similar a la de la resta de

Wikipedies de llengües d'extensió més petita. Per tal d'estudiar amb més profunditat els factors que expliquen la dimensió de les altres llengües naturals a la Wikipedia, hem fet una anàlisi amb només aquests *altres* 21 casos.²⁵

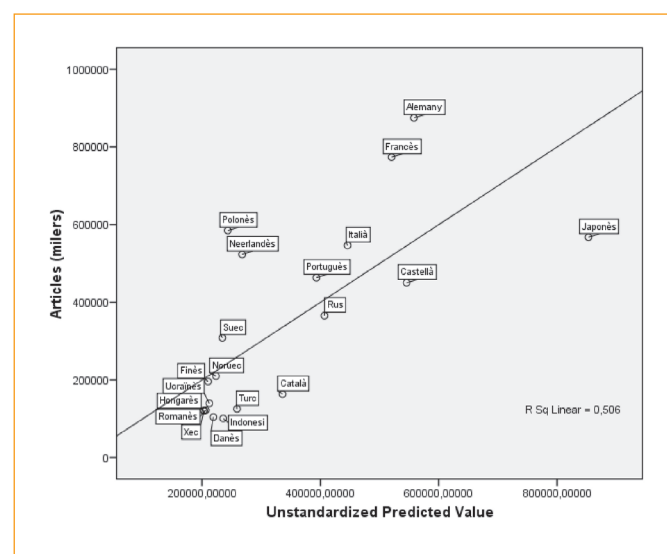
La primera anàlisi mostra que els factors demolingüístics perden rellevància en l'explicació de la dimensió d'aquestes Wikipedies i que els factors econòmics s'erigeixen com a principals valors predictius. El PIB prediria fins al 43,7% de la variació en el nombre d'articles. Si ens fixem en el gràfic 6, podem observar que hi ha una sèrie de llengües força distants de la recta de regressió. D'una banda, hi ha les que se situen per la banda inferior de la línia, entre les quals es destaca el japonès, al qual corresponen, segons la predicció, molts més articles dels que haurien de tenir pel seu PIB. S'ha de ser més prudent, però, en el cas del català, que també es destaca per sota de la recta de regressió. Tal com s'indica a la nota 21, no és fàcil la comparació del PIB dels territoris de llengua catalana amb les dades internacionals i s'utilitzen els resultats de tot l'Estat espanyol. A la banda oposada hi ha l'alemany, el francès i el polonès, però també el neerlandès, el portuguès i el castellà, que tenen un nombre d'articles molt superior al que s'ha predit pel PIB de cadascuna de les llengües.

Una vegada explorat el principal paper que hi tenen els factors econòmics, hem buscat una segona variable que optimitze el model. Aquests factors són sociolingüístics, amb el nombre de parlants de cadascuna de les llengües. Però el model final té un

Gràfic 6. Model 1. Representació mitjançant un núvol de punts de la dimensió de la Wikipedia segons el PIB. 21 principals llengües naturals a la Wikipedia (anglès i xinès exclosos). $R^2 = 0,437$.



Gràfic 7. Model 2. Representació mitjançant un núvol de punts de la dimensió de la Wikipedia segons la puntuació en la regressió (PIB i total de parlants). 21 principals llengües naturals a la Wikipedia (anglès i xinès exclosos). $R^2 = 0,506$.



25. Les 21 llengües estudiades en aquest model són l'alemany, el francès, el polonès, el japonès, l'italià, el neerlandès, el portuguès, el castellà, el rus, el suec, el noruec, el finès, el català, l'ucraïnès, el turc, el romanès, el txec, l'hongarès, l'eslovac, el danès i l'indonesi.



<http://digithum.uoc.edu>

La dimensió de les llengües a la Wikipedia i la seua relació...

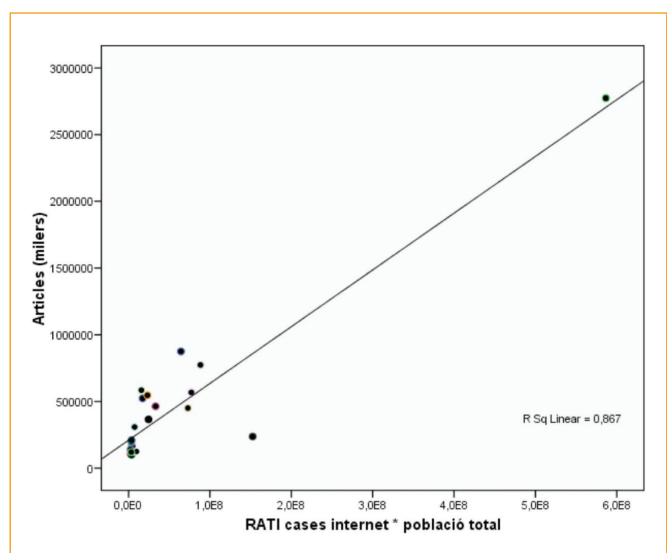
increment molt baix de capacitat explicativa. Arriba només ja al 50,6% d'explicació de la variació. Com es pot veure al gràfic 7, els únics canvis realment importants són el fet que si bé en el model anterior el castellà tenia un nombre d'articles molt superior al nombre predit pel seu PIB, ara passa a tenir pocs articles segons el nombre predit per la combinació de PIB i de nombre total de parlants.

Un cop estructurat aquest model que explica la meitat de la variació del nombre d'articles segons factors de caire econòmic i sociolingüístic, cap variable respon amb prou solidesa per millorar el model.

Conclusions

Les anàlisis han mostrat que la dimensió de les principals llengües a la Wikipedia és marcada pel comportament diferenciat de quatre tipologies de llengües: l'anglès, llengua hipercentral; el xinès, llengua amb un nombre més elevat de parlants inicials; la resta de llengües naturals i les dues llengües construïdes, aquestes darreres excloses de l'anàlisi. També hem vist les correlacions molt altes entre la major part de les variables que s'han estudiat, siguen de tipus social, sociolingüístic, econòmic, tecnològic, educatiu o demogràfic, i que, malgrat que algunes mostren una correlació més gran que d'altres amb la dimensió de la llengua a la Wikipedia, no és difícil l'intercanvi entre elles.

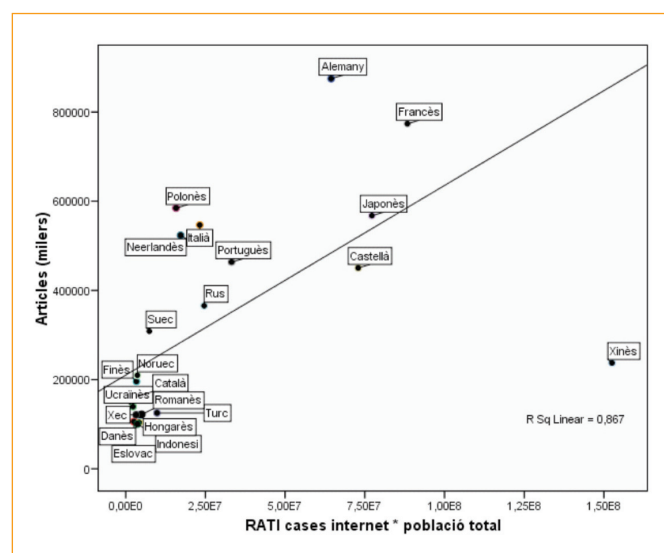
Gràfic 8. Representació mitjançant un núvol de punts de la dimensió de la Wikipedia segons la dimensió de la comunitat lingüística digital. 23 principals llengües naturals a la Wikipedia.



Per això, per tal de millorar la idea que hi ha de rerefons en els resultats, hem proposat el concepte de *comunitat lingüística digital*. La idea que jau darrere d'aquest concepte és simplificar les anàlisis. És a dir, si l'amplitud de la comunitat lingüística d'una llengua és cabdal per a explicar-ne la dimensió a la Wikipedia, però no és suficient i ha de ser complementada amb altres variables de tipologia diferent, el matís aportat per la *comunitat lingüística digital* enfortiria la capacitat explicativa del primer factor. Aquesta variable simplement pondera el nombre de parlants d'una llengua segons la difusió d'internet entre la comunitat lingüística corresponent. Indica, per exemple, en el cas del xinès, que tot i que té un total de 930 milions de parlants, només el 16,4% de les llars tenen internet, amb la qual cosa la comunitat lingüística digital que potencialment pot editar a la Wikipedia es redueix a 153 milions d'usuaris. Per contra, la comunitat lingüística de 906 milions de parlants de l'anglès, amb unes taxes de penetració d'internet del 64,7%, formen una comunitat lingüística digital de 587 milions d'usuaris, la més gran de totes amb diferència.²⁶

En el gràfic 8 veiem la disposició de les principals llengües naturals de la Wikipedia segons la dimensió de la seua comunitat lingüística digital. El model s'ajusta d'una manera molt notable, amb una R^2 que significa un 86,7% d'ajustament. S'observa com a la banda esquerra superior se situa l'anglès, amb la comunitat lingüística digital

Gràfic 9. Representació mitjançant un núvol de punts de la dimensió de la Wikipedia segons la dimensió de la comunitat lingüística digital. 23 principals llengües naturals a la Wikipedia. Visió ampliada a les 22 principals llengües naturals, excloent-ne l'anglès.



26. Les comunitats lingüístiques d'usuaris d'internet són (en milions d'usuaris): anglès 586, alemany 64, francès 88, polonès 15, japonès 77, italià 23, neerlandès 17, portuguès 33, castellà 72, rus 24, suec 7, xinès 152, noruec 3, finès 3, català 4, ucraïnès 2, turc 9, romanès 5, txec 3, hongarès 4, volapük 0, esperanto 0, eslovac 2, danès 4 i indonesi 3.



<http://dighum.uoc.edu>

La dimensió de les llengües a la Wikipedia i la seua relació...

més gran i el nombre d'articles més elevat, mentre que la resta de llengües, inclòs el xinès, tenen un nombre d'articles inferior, conseqüència de les seues comunitats lingüístiques d'usuaris més petites.

Al gràfic 9 es pot observar una ampliació del marge inferior esquerra del gràfic 8, amb la corresponent denominació de cadascuna de les llengües. S'hi observa que entre aquestes altres llengües la distància respecte a la recta de regressió és molt més gran. De fet, si fem la mateixa regressió entre la comunitat lingüística digital i la dimensió de la Wikipedia, però n'excloem l'anglès i el xinès, el model baixa l'ajust fins al 65,3%. I si només n'excloem l'anglès i hi deixem el xinès, no arriba al 25,5%. Per contra, si l'exclusió és la del xinès i hi deixem l'anglès, arriba al 92,7%.

Per tant, observem que la definició d'una *comunitat lingüística digital* simplifica l'anàlisi de bon començament i el fa més raonable. No obstant això, l'anàlisi ha de millorar la capacitat explicativa, sobretot pel que fa al comportament de les llengües mitjanes i grans. L'exclusió dels casos atípics, com l'anglès i el xinès, pot ajudar en el futur a aclarir els principals factors que situen les llengües a la Wikipedia.

Referències bibliogràfiques

1 Cog. «Don't say "search engine" – say "Google"» [article en línia]. [Data de consulta: 24 de juny de 2009].

<<http://1cog.com/search-engine-statistics.html>>

Alexa Internet. «Top 500 Global Sites» [article en línia]. [Data de consulta: 24 de juny de 2009].

<<http://www.alexa.com/topsites>>

<http://ca.wikipedia.org/wiki/Alexa_Internet>

Alexa Internet. «Five-year Traffic Statistics for Wikipedia.org» [article en línia]. [Data de consulta: 24 de juny de 2009].

<<http://www.alexa.com/siteinfo/wikipedia.org?range=5y&size=large&y=t>>

<http://ca.wikipedia.org/wiki/Alexa_Internet>

BELZUNCES, Marc (2008). «Les 40 llengües de Google, o la importància de tenir el navegador en català» [article en línia]. *De l'Holocè estant. Bloc personal d'en Marc Belzunces*. [Data de consulta: 23 de juny de 2009].

<<http://blocs.mesvilaweb.cat/node/view/id/108618>>

Ant's Eye View. «The 90-9-1 Principle» [article en línia]. [Data de consulta: 24 de juny de 2009].

<<http://www.antseyeview.com/90-9-1-principle/>>

CALVET, Louis-Jean (2006). *Towards an Ecology of World Languages*. Cambridge: Polity Press.

CASTELLS, Manuel (2003). *L'era de la informació: economia, societat i cultura*. Barcelona: Editorial UOC.

DANS, Enrique (2009). *Encarta, un caso interesante* [article en línia]. [Data de consulta: 24 de juny de 2009].

<<http://www.enriquedans.com/2009/04/encarta-un-caso-interesante.html>>

Enciclopedia libre universal en Español.

<<http://enciclopedia.us.es/>>

Ethnologue. *Fourteenth Edition. Language name index* [en línia]. [Data de consulta: juny de 2009].

<http://www.ethnologue.com/14/language_index.asp>

Linguamón - Casa de les Llengües. *Les llengües del món* [en línia]. [Data de consulta: juny de 2009].

<http://www10.gencat.cat/casa_llengues/AppJava/ca/diversitat/diversitat/llengues_mon.jsp>

MANRIQUE, Rosendo (2009). «Models de negoci antiquats davant la web 2.0?» [article en línia]. *Livejournal*. [Data de consulta: 24 de juny de 2009].

<<http://skaps.livejournal.com/158348.html>>

Organització de les Nacions Unides (2008). *World Population Prospects* [en línia]. [Data de consulta: juny de 2009].

<<http://esa.un.org/unpp/>>

ORTEGA SOTO, José Felipe (2009). *Wikipedia: A quantitative analysis*. Tesi doctoral. Madrid.

<http://librosoft.es/Members/jfelipe/thesis-wkp-quantanalysis/at_download/file>

PÉREZ-LANZAC, Carmen (2009). «Debemos fiarnos de la Wikipedia?» [article en línia]. *El País.com*. [Data de consulta: 10 de juny de 2009].

<http://www.elpais.com/articulo/sociedad/Debemos/fiarnos/Wikipedia/elpepusoc/20090610elpepusoc_1/Tes>

POZO, Òscar del (2008). «Vídeo dels parlaments de la festa del 10è aniversari de Softcatalà» [vídeo en línia; a partir del minut 39]. *De l'Holocè estant. Bloc personal d'en Marc Belzunces*. [Data de consulta: 23 de juny de 2009].

<<http://blocs.mesvilaweb.cat/node/view/id/110400>>

PROGRAMA DE LES NACIONS UNIDES PER AL DESENVOLUPAMENT (PNUD) (2008). *Informe sobre el desenvolupament humà 2007-2008* [en línia].

<http://hdr.undp.org/en/media/HDR_20072008_SP_Complete.pdf>

Racó Català (6 d'octubre de 2008). «La importància de fer notar a Google que es té el navegador configurat en català» [article en línia]. [Data de consulta: 23 de juny de 2009].

<<http://www.racocatala.cat/noticia/17383/importancia-fer-notar-google-te-navegador-configurat-catala>>

Site analytic. «Site comparison of google.com, yahoo.com, msn.com» [article en línia]. [Data de consulta: 24 de juny de 2009].

<<http://siteanalytics.compete.com/google.com+yahoo.com+msn.com/?metric=uv>>

SWAAN, Abram de (2001). *Words of the World. The Global Language System*. Cambridge: Polity Press.

TANCER, Bill (2007). «Look Who's Using Wikipedia» [article en línia]. *Time*. [Data de consulta: 1 de maig de 2007].

<<http://www.time.com/time/business/article/0,8599,1595184,00.html>>

<http://ca.wikipedia.org/wiki/Time_%28magazine%29>



<http://digithum.uoc.edu>

La dimensió de les llengües a la Wikipedia i la seua relació...

UNIÓ INTERNACIONAL DE TELECOMUNICACIONS (2009). «Measuring the information society» [article en línia]. [Data de consulta: juny de 2009].

<http://www.itu.int/ITU-D/ict/publications/idi/2009/material/IDI2009_w5.pdf>

VAN DIJK, Ziko (2009). «Wikipedia and lesser-resourced languages». *Language Problems and Language Planning*. Vol. 33, núm. 3, pàg. 234-255.

Viquipèdia. «Web 2.0» [article en línia]. [Data de consulta: 24 de juny de 2009].

<http://ca.wikipedia.org/wiki/Web_2.0>

Wikimedia. «Research» [article en línia]. [Data de consulta: 10 de maig de 2011].

<<http://meta.wikimedia.org/wiki/Research>>

Wikimedia. «Wiki Research Bibliography» [article en línia]. [Data de consulta: 10 de maig de 2011].

<http://meta.wikimedia.org/wiki/Wiki_Research_Bibliography>

Wikipatterns. «90-9-1 Theory» [article en línia]. [Data de consulta: 24 de juny de 2009]

<<http://www.wikipatterns.com/display/wikipatterns/90-9-1+Theory>>

Wikipedia. «Wikipedia: Academic studies of Wikipedia» [article en línia]. [Data de consulta: 10 de maig de 2011].

<http://en.wikipedia.org/wiki/Wikipedia:Wikipedia_in_academic_studies>

WOODSON, Alex (2007). «Wikipedia remains go-to site for online news» [article en línia]. *Reuters*. [Data de consulta: 24 de juny de 2009].

<<http://www.reuters.com/article/2007/07/08/us-media-wikipedia-idUSN0819429120070708>>

<<http://ca.wikipedia.org/wiki/Reuters>>

ZLATIĆ, Vinko; BOŽIČEVIĆ, Miran; ŠTEFANČIĆ, Hrvoje; Domazet, Mladen (2006, 3 de juliol). «Wikipedias: Collaborative web-based encyclopedias as complex networks». *Physical Review*.

<http://arxiv.org/PS_cache/physics/pdf/0602/0602149v3.pdf>

Agraïments

Aquest estudi no hauria estat possible sense la idea i l'emprenedoria de Marina Nogué,²⁷ investigadora del Centre Universitari de Sociolingüística i Comunicació (CUSC) de la Universitat de Barcelona. Els autors de l'article ens sentim deutors de les seues forces davant els grans reptes. A ella li devem molt més que una idea. F. Xavier Vila i Moreno²⁸ ens animà a tirar endavant el projecte, a donar-hi la rellevància que es mereix i oferir al nostre projecte el ple suport des de la direcció de la Xarxa CRUSCAT - IEC. Jaume Llopis, assessor estadístic, ens donà el suport necessari per a pensar la relació entre la Wikipedia i les variables socials des de diferents perspectives. Samuel Gimeno,²⁹ Daniel Torres Burriel,³⁰ Toni Hermoso,³¹ Juan Pablo Martínez,³² Manuel Trujillo Berge³³ i els participants de la Viquitrobada 2009³⁴ ens han ajudat a navegar per les tripes de la Wikipedia, traure-hi la informació que necessitàvem i entendre el paper de les llengües mitjançes i menudes a internet.

CITACIÓ RECOMANADA

PELLEJERO, Borja; SOROLLA, Natxo; NOGUÉ, Marina (2011). «La dimensió de les llengües a la Wikipedia i la seua relació amb els elements socials» [article en línia]. *Digithum*. Núm 13, pàg. 37-49. UOC. [Data de consulta: dd/mm/aa].

<<http://digithum.uoc.edu/ojs/index.php/digithum/article/view/n13-pellejero-sorolla-nogue/n13-pellejero-sorolla-nogue-cat>>

ISSN 1575-2275



Els textos publicats en aquesta revista estan subjectes –llevat que s'indiqui el contrari– a una llicència de Reconeixement 3.0 Espanya de Creative Commons. Podeu copiar-los, distribuir-los, comunicar-los públicament i fer-ne obres derivades sempre que reconegueu els crèdits de les obres (autoria, nom de la revista, institució editora) de la manera especificada pels autors o per la revista. La llicència completa es pot consultar a <http://creativecommons.org/licenses/by/3.0/es/deed.ca>.

27. Vegeu: <<http://lullabyinblackandwhite.wordpress.com/>>.

28. Vegeu: <<http://fxvila.wordpress.com/>>.

29. Vegeu: <<http://www.enver555.net/blog/>>.

30. Vegeu: <<http://www.torresburriel.com/>>.

31. Vegeu: <<http://www.cau.cat/blog/qui>>.

32. Vegeu: <<http://diec.unizar.es/~jpmart/personal/>>.

33. Vegeu: <<http://manueltrujilloberges.blogspot.com/>>.

34. Vegeu: <http://ca.wikipedia.org/wiki/Viquiprojecte:Tercera_trobada_i_primer_congr%C3%A9s_viquipedista_2009>.

**Borja Pellejero**

sarenalet@gmail.com

Col·laborador de la Xarxa CRUSCAT – IEC

Borja Pellejero és llicenciat en Sociologia per la Universitat de Barcelona l'any 2009. Ha treballat a la Xarxa CRUSCAT amb un contracte de pràctiques i al Centre d'Estudis Locals d'Alaior.

Natxo Sorolla Vidal

nsorolla@iec.cat

Investigador de la Xarxa CRUSCAT – IEC

Natxo Sorolla és sociòleg i fa la seua tesi doctoral a la Universitat de Barcelona sobre l'anàlisi de les xarxes socials, el manteniment i la substitució lingüística en àrees rurals. Ha participat en diferents publicacions sobre demolingüística, sociolingüística i xarxes socials als territoris de llengua catalana, entre les quals es destaquen el volum d'anàlisi d'enquestes *Llengua i societat als territoris de parla catalana a l'inici del segle XXI* i l'*Informe sobre la situació de la llengua catalana* en les seues dues darreres edicions (2005-07 i 2007-08). És administrador de *Lafranja.net* (<<http://www.lafranja.net>>) i manté un blog personal sobre el seu objecte d'estudi (<<http://xarxes.wordpress.com/>>).

Institut d'Estudis Catalans

C. del Carme, 47

08001 Barcelona

Marina Nogué Pich

Membre del Centre Universitari de Sociolingüística i Comunicació (UB)

Marina Nogué ha estat col·laboradora del Termcat i coautora de la publicació *Estudis d'implantació terminològica: una aproximació en l'àmbit dels esports*. També ha col·laborat amb el Departament de Filologia Catalana de la Universitat de Barcelona i ha estat guardonada amb el premi CIRIT per a fomentar l'esperit científic del jovent, concedit per la Comissió Interdepartamental de Recerca i Innovació Tecnològica de la Generalitat de Catalunya.